

科技资源站点网络影响力的评价

周琼琼^{1,2} 张 辉³ 华青松⁴

(1. 国家科技基础条件平台中心, 北京 100862; 2. 西南交通大学 经济管理学院, 四川成都 610031;
3. 北京航空航天大学, 北京 100191; 4. 大连华锐重工集团, 辽宁大连 116013)

摘 要: 利用主成分分析法, 对国家科技基础条件平台形成的20个科技资源站点的7个网络流量指标进行综合评价, 使用MATLAB软件进行辅助计算和分析, 提出用3个主成分的标准值来衡量这些科技资源站点的网络影响力, 从而反映科技资源站点资源信息化建设和共享服务水平, 最后提出增强科技资源站点影响力的相关建议。

关键词: 主成分分析; 网站影响力; 评价指标; 科技资源网站

中图分类号: G206

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2013.02.015

Evaluation of R&D Resources Network Influence Degree

Zhou Qiongqiong^{1,2}, Zhang Hui³, Hua Qingsong⁴

(1. National Science and Technology Infrastructure Center, Ministry of Science and Technology of China, Beijing 100862; 2. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031; 3. Beihang University, Beijing 100191; 4. Dalian Huarui Hemry Industry Co.Ltd., Dalian 116013)

Abstract: In this paper, the author makes general evaluation on seven evaluation indexes of 20 R&D resources networks with principal component analysis. Through MATLAB for calculation and analysis, the author puts forward three principal components as the measurement of R&D resource network influence level and gets the orders of them, which shows the different types of R&D resource integration development and sharing service level. At last, the author presents relevant suggestions to improve R&D resources networks performance.

Keywords: principal component analysis, network influence degree, evaluation index, R&D resources network

1 引言

根据“整合、共享、完善、提高”的指导思想, 自2004年以来, 国家科技基础条件平台开展了大量科技资源的收集整理和整合共享工作, 并在此基础上, 充分运用信息、网络等现代技术手段^[1], 面向全社会共享和开放科技资源, 为科技工作者提供了大量实物资源和信息资源服务。截至目前, 中央财政累计投入平台专项经费近30亿元, 搭建了由研究实验基地和大型科学仪器设备、自然资源、科学数据、科技文献等六大共享平台为主体框架的国家科技基础条件平台, 初步形成了跨部门、跨区域、多层次的资源整合共享体系, 并建立了国

家科技基础条件平台门户系统, 形成了逻辑上统一、物理上分布的科技资源服务架构。

此外, 各资源平台为提高科技资源的使用效率和综合利用水平, 相继建立了各自的科技资源站点和完备的网络服务系统。这些科技资源站点已经成为社会公众, 特别是科技工作者获取平台科技资源信息和共享服务的主要渠道, 产生了广泛的社会关注度和网络影响力。

网络影响力评价是一种基于客观信息计量和效用分析的评价方法, 能够客观地反映信息资源管理和利用状况^[2]。在信息资源管理研究领域, 基于客观信息计量的网络信息资源分析和评价研究在理论方法上的发展十分迅速。网络信息计量学诞生于

第一作者简介: 周琼琼(1981—), 女, 助理研究员, 研究方向: 科技管理。

收稿日期: 2013年1月12日。

1997年,它以探索网络信息资源的生产、传播、分布和利用规律为基本任务,逐步形成了以网络链接分析为基础的理论和方法体系^[3]。目前,在国内外网络影响力评价的内容较为广泛,也逐渐受到重视,成为理论研究和应用研究的热点内容之一。国内外知名机构不约而同地把“影响力”作为网络信息资源调查和评价的重要方面,如学术论文、期刊影响力评价是衡量学术水平的重要标准,政府网站的网络影响力已经成为衡量电子政务建设绩效的重要指标。如全球标准互联网用户调查和分析权威机构尼尔森通过对网络发展趋势的多年追踪,根据对英国用户的调查结果,评选出了2003年最具影响力的10大网站^[4]。中国互联网络信息中心一直开展中国互联网络信息资源的调查和统计,并发布中国互联网络发展状况和中国互联网络信息资源数量的统计年度报告。自1998年起,中国互联网络信息中心先后5次发布了《中国互联网络网站影响力调查报告》^[5]。

总体来看,国内关于网络影响力评价的研究一般较注重理论研究,如评价指标体系的建立、评价方法的研究等。而国外则在进行理论研究的同时,则更侧重于应用和对实践的指导。例如,他们对于评价指标的获取、统计分析和验证研究较为深入具体。国外许多研究是通过绩效评估来验证信息资源质量,进而实现对信息质量的管理,而不是单纯地用某个指标体系进行定性分析或定量评价。但总的来说,基于科技资源站点本身的网络影响力评价无论从研究理论、实证分析等都不多见。

本文所提出的对科技资源站点网络影响力评价主要是通过对科技资源站点的建设水平与综合访问情况等进行了客观评估,以衡量各科技平台在资源建设和共享服务方面所取得的成效,从而为提高科技平台的运行服务效果提供参考依据。需要明确的是,科技资源站点影响力评价只是科技基础条件平台绩效评价的一部分内容,是平台绩效平台的补充和完善,不可能完全覆盖平台绩效的全部功能范畴。另外,由于影响力本身就是带有信息化色彩,因此,不同于对资源本身的评价,对科技资源站点网络影响力评价只能反映对网站本身受关注和使用的程度,还有很多其他方面需要进一步探究。

本文构建了科技资源站点网络影响力评价指标体系,利用主成分分析方法^[6],对科技基础条件平台门户系统所采集到的各类资源站点数据进行对比

分析,以期科技管理部门对科技资源平台的资源建设和运行服务方面的绩效评价提供决策参考。

2 评价指标与样本数据

本文评价的主体是各类资源建设的科技资源站点。资源站点是国家科技基础条件平台开展运行服务的重要形式和有效载体。从这个意义上讲,科技资源站点的网络影响力从一个侧面反映了各类科技资源平台的资源信息化建设水平和共享服务能力。为确保采样的准确性和可对比性,本文以国家科技基础条件平台门户下设的运行相对稳定的20个主要资源科技资源站点为研究对象,参考国际上通用的网络流量指标^[7-8],遴选了7个具有代表性的网络流量指标作为影响力评价的主要内容。这些指标能够在一定程度上反映科技资源的内容丰富度、信息化水平和用户使用状况,从而间接地体现了各类科技资源平台的社会认可程度和影响力水平。统计数据来源于中国科技资源共享网的评估监测系统。该评估监测系统采用网站日志统计分析工具对各类指标进行采集和统计,采样时间周期为一个月。科技资源站点网络影响力评价指标有如下几项。

(1)访问量 X_1 :在一定时间内访问某一网站的请求次数,根据访问用户的IP地址进行统计。同一个IP在不同时间、不同浏览器进行访问视为不同的访问次数。

(2)访问人数 X_2 :在一定有效时间内访问服务器页面的不同用户数量,系统可以根据不同的IP地址或其他参数识别不同用户。

(3)访问页面数 X_3 :对网站不同页面被点击访问的统计,类同于用户的点击数。用户的一次访问可带来多次的访问页面统计,等同于用户访问某一网站时查看该网站的网页数。

(4)访问者参与指数 X_4 :该指标反映了每个访问者对某一网站所访问的平均次数,主要用于反映访客回访意愿的强烈程度,其计算方法为总访问数/独立访问人数。如果网站具有明确的目标用户群,则该指数将会远远大于1;如果网站的回访者较少,该指数将趋近于1。该指数的高低取决于网站的定位,例如内容型以及商业型的网站都期望该指数较高,而投诉类或客户服务类网站则希望该指数尽量接近于1。

(5)黏着度 X_5 :用户每人访问网站时平均浏览的页面数,是反映平台资源信息的丰富程度和对

用户吸引程度的重要指标,计算方法为访问页面数/访问量。网站对于用户的黏着主要来自两方面,内容黏着或者是用户黏着。内容黏着是指网站由于信息资讯的丰富或内容的高度专业化而对用户产生的吸引;用户黏着是指用户根据其网络社交关系而对网站产生的使用依赖。

(6)响应时间 X_6 :用户访问一个网站时该网站响应访问所需的时间,是衡量网站反应快慢和网站友好性的指标。

(7)失效率 X_7 :用户在访问网站时,遇到的网站返回出错信息的次数与总访问请求次数的比率,用于衡量该网站的建设质量和服务质量。网站的出错信息是通过日志文件中的访问响应状态参数获取的,本文中样本统计的错误信息包括HTTP 404及HTTP 500。

从网络运营的角度还可以采用网站转换率、回访者比率、积极访问者比率、回弹率、浏览用户比率^[9-10]等内容指标,来把握用户动向和网站的实际状况。

鉴于中国科技资源共享网评估监测系统统计指标内容所限,在剔除个别指标统计数据不完全和异常的科技资源站点后,本文根据平台门户系统某年度的监测数据(表1),通过主成分分析方法进行了数据分析。样本相关系数矩阵如表2所示。为保证各指标数据的正相关性,将平均响应时间进行了取负值处理,对失效率进行了取倒数处理。

3 主成分值排序

从表2可以看出,用户访问量与访问人数的相关系数达到了0.950138,说明这两个指标在衡量科技

表1 某年度科技资源站点样本数据

序号	资源站点	访问量 (次)	访问 人数(位)	访问页面数 (个)	访问者参与 指数(次/人)	黏着度 (页/次)	响应 时间(ms)	失效率 (%)
1	国家农作物种质资源平台	204312	66808	1043876	3.058197	5.109225	1291	20.67
2	国家微生物资源平台	157256	39864	2161418	3.944812	13.74458	5346	11.47
3	国家标准物质资源共享平台	80970	21437	1245528	3.777114	15.38259	1842	27.15
4	标本资源共享平台	507846	167055	7547592	3.039993	14.86197	2436	16.82
5	家养动物种质资源平台	80875	25932	269372	3.118734	3.33072	3201	28.70
6	水产种质资源平台	9246	6534	48957	1.41506	5.294938	1084	35.22
7	林业科学数据平台	240724	74872	3991366	3.215141	16.58067	1353	39.95
8	地球系统科学数据共享平台	636810	313586	11295748	2.030735	17.73802	777	36.42
9	人口与健康科学数据共享平台	720624	261687	5685135	2.753763	7.889184	915	15.89
10	农业科学数据共享中心	45901	17193	306936	2.669749	6.686913	9280	18.13
11	气象科学数据共享中心	444589	144280	7895911	3.081432	17.76002	272	23.38
12	地震科学数据共享中心	403246	124190	3709351	3.247009	9.19873	2309	12.51
13	国家科技图书文献中心	871935	283399	213393783	3.076705	244.7359	1210	8.63
14	国家标准文献共享服务平台	404116	144274	6434491	2.801031	15.92239	288	9.49
15	国家生态系统观测研究网络	46524	13204	389858	3.523478	8.379718	137	12.13
16	国家大型科学仪器中心	8630	6927	34884	1.24585	4.042178	2317	14.13
17	国家计量基标准(物理部分) 资源共享基地	78086	23196	252488	3.366356	3.233461	95	8.63
18	国家计量基标准(化学部分) 资源共享平台	120098	46343	364663	2.591502	3.036379	1761	12.23
19	中国应急分析测试平台	268244	180092	2427484	1.489483	9.049537	7880	12.23
20	中国检测资源平台	416280	121424	45648165	3.428317	109.6574	559	5.69

表2 科技资源站点影响力指标相关系数阵

指标	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1	0.950138	0.609726	0.07188	0.595559	-0.10268	0.259236
X_2		1	0.491532	-0.12543	0.464758	-0.06219	0.098535
X_3			1	0.107098	0.9769	0.192513	0.347566
X_4				1	0.164222	-0.22068	0.21331
X_5					1	0.211795	0.376439
X_6						1	-0.02163
X_7							1

资源站点网络影响力方面体现的价值相当。从网络计量学的角度来看，用户访问量可以充分体现一个科技资源站点的内容丰富度。另外，访问页面数和黏着度的相关系数为0.9769，说明访问页面数在很大程度上体现了资源站点对于用户的吸引程度。从表3看出，特征值大于1的主成分有3个，且他们的累计贡献率大于80%，故可用前3个主成分变量来代替原来的7个变量，主成分对应的特征向量列于表4。

根据以上分析，可以计算出各主成分的值。

$$\text{PINZ}_1=0.490952 \times X_1-0.07167 \times X_2+0.370373 \times X_3+0.001317 \times X_4-0.28973 \times X_5-0.00793 \times X_6-0.72983 \times X_7$$

$$\text{PINZ}_2=0.43485 \times X_1-0.25781 \times X_2+0.451441 \times X_3+0.017756 \times X_4-0.3059 \times X_5+0.001207 \times X_6+0.668394 \times X_7$$

$$\text{PINZ}_3=0.501397 \times X_1-0.00966 \times X_2-0.26233 \times X_3-0.15953 \times X_4+0.406025 \times X_5+0.698632 \times X_6+0.036047 \times X_7$$

主成分标准值PINZ是PINZ₁、PINZ₂与PINZ₃加权平均值，相应的“权”分别是前两个特征值的贡献率，计算方法为PINZ=0.457416*PINZ₁+0.1867*PINZ₂+0.177468*PINZ₃。需要指出的是，由于主成分标准值的数据正负只是反映了各资源站点与平均水平的相对情况，将零点作为平均水平，得分为正表明处于平均水平之上，得分为负表示处于平均水平之下。由于所有指标都是正相关性指标，因此得分越高表明科技资源站点的网

络影响力越高。为更好地区分各资源站点评价情况，我们将3个主成分PINZ进行线性变换得到PINZ’=50PINZ+80，PINZ’称为主成分变换值。计算结果如表5所示。

从主成分特征值可以看出，访问量、访问页面数、访问者参与指数、黏着度和失效率都是网络影响力评价的关键因子。其中，访问量、访问页面数可以归纳为是从用户使用角度体现科技资源站点影响力的指标，访问者参与指数、黏着度和失效率可归纳为是从用户认可角度体现科技资源站点影响力的指标。在第一个主成分公式中，访问量、访问页面数和黏着度具有较大的正值，而在失效率具有较大的负值，其含义是如果科技资源站点的访问量、访问页面数和粘着度越高，而且错误率越低，则其内容信息的丰富度以及对用户的吸引程度越好，网络影响力越大。按照第一主成分之值排序结果显示，国家科技图书文献中心明显居于首位，中国检测资源平台和地球系统科学数据共享平台次之。这与访问量、访问页面数的排名相对一致，但与其中的任何一个指标并不总保持完全一致。这说明资源站点的影响力评价不能简单地用其中任何一个指标完全代表，即便是采用最有代表性的访问量X₁也不行，应考虑访问者参与指数与粘着度等多项指标的综合情况而进行综合评价。

表3 主成分与特征值

主成分	特征值	贡献率（%）	累计贡献率（%）
Z ₁	3.201911	0.457416	0.457416
Z ₂	1.306898	0.1867	0.644116
Z ₃	1.242277	0.177468	0.821584
Z ₄	0.717765	0.102538	0.924121
Z ₅	0.492363	0.070338	0.994459
Z ₆	0.020435	0.002919	0.997378
Z ₇	0.018352	0.002622	1

表4 前3个主成分对应的特征向量

指标	主成分		
	Z ₁	Z ₂	Z ₃
X ₁	0.490952	0.43485	0.501397
X ₂	-0.07167	-0.25781	-0.00966
X ₃	0.370373	0.451441	-0.26233
X ₄	0.001317	0.017756	-0.15953
X ₅	-0.28973	-0.3059	0.406025
X ₆	-0.00793	0.001207	0.698632
X ₇	-0.72983	0.668394	0.036047

表5 某年度科技资源站点主成分值排序

资源类型	第一主成分	第二主成分	第三主成份	主成份变换值	排序
国家科技图书文献中心	6.3214	-0.1687	-1.3938	211	1
人口与健康科学数据共享平台	1.0475	0.5185	3.4665	140	2
中国检测资源平台	1.7359	0.8924	-0.8485	121	3
地球系统科学数据共享平台	1.4134	-1.7273	1.3061	108	4
标本资源共享平台	0.4819	-0.0962	0.8125	97	5
气象科学数据共享中心	0.4532	0.2739	0.3961	96	6
地震科学数据共享中心	0.4801	0.7717	-0.28	96	7
国家微生物资源平台	-0.5854	2.1201	0.209	88	8
国家标准文献共享服务平台	0.3767	-0.3527	-0.2117	83	9
国家计量基标准（化学部分） 资源共享平台	-0.961	1.2002	1.0751	79	10
林业科学数据平台	-0.6729	-0.3388	-0.1905	60	11
国家计量基标准（物理部分） 资源共享基地	-0.957	0.8905	-0.767	60	12
国家农作物种质资源平台	-0.7736	-0.0896	-0.2238	60	13
国家生态系统观测研究网络	-1.0971	1.0358	-0.6668	59	14
国家标准物质资源共享平台	-1.1545	0.7111	-0.4944	56	15
家养动物种质资源平台	-1.3556	0.283	0.024	52	16
中国应急分析测试平台	-0.5921	-2.6751	0.6833	48	17
农业科学数据共享中心	-1.3581	-0.5488	-0.8799	36	18
国家大型科学仪器中心	-1.0638	-0.9393	-1.4772	34	19
水产种质资源平台	-1.7391	-1.7609	-0.5392	19	20

4 评价结果分析

总体而言, 20个科技资源站点已初步形成了一定数量的用户群, 并提供了内容丰富的科技资源信息服务, 但由于各类科技资源站点的类型不同, 科技资源站点的主成分变化值差别较大, 科技资源所具备的特有属性在一定程度上影响了资源的综合利用效率和受众规模。主成分变化值在200~300之间的资源站点有1个, 主成分变化值在100~200之间的科技资源站点有3个, 主成分变化值在0~100之间的科技资源站点较多, 共有16个且分布较均匀。其中, 0~50主成分变化值在0~100之间的有5个, 呈现出两端分布少、中间分布较多的情况。其中, 国家科技图书文献中心的访问页面数达到亿万次, 在20个资源站点遥遥领先。而访问页面数达到千万级的资源站点有中国检测资源平台和地球系统科学数据共享平台2个。访问页面数达到百万级以上的资源站点有13个。说明这些科技资源站点的网络影响力已经达到一定水平, 受到了社会的广泛关注。

从用户参与度角度看, 国家微生物资源平台、国家标准物质资源平台和国家生态系统观测研究网络资源站点的访问者参与指数较高, 说明这些平台已经形成了一批较稳定的特定用户群, 且资源对用户具有较大的吸引力; 全国科学仪器网、水产种质

资源平台和中国应急分析测试平台的访问者参与指数较低, 说明这些资源站点对用户吸引力不大, 在内容建设和用户互动性建设方面需进一步加强。

从用户黏着度角度看, 国家科技图书文献中心的黏着度接近245, 中国检测资源平台的用户黏着度接近110, 在20个资源站点中明显居于前列, 这说明图书文献资源和检测资源的重复使用率相对较高; 相比而言, 家养动物种质资源、水产种质资源、国家计量基标准（物理部分、化学部分）资源站点的用户黏着度相对较低, 这说明用户对于该类型资源信息的反复使用率不高, 这些资源站点应加强对资源的深层数据挖掘和数据产品加工、推广等方面工作。

从资源站点响应时间看, 这些科技资源站点相差较为悬殊。响应时间直接决定了用户享受资源信息服务的体验, 对于提高科技资源站点的访问量、访问者参与度和黏着度有着潜移默化的影响。从第二主成分的变量值中可以看出, 响应时间从一个侧面决定了科技资源站点的网络影响力水平, 因此在进行科技资源服务网站的建设中, 网站结构设计应清晰、合理, 应在细节上多关注用户的实际体验。

综上所述, 科技资源站点的网络影响力与科技资源属性、规模、用途和受众面等因素息息相关。以科技图书文献中心为例, 主成分变化值达

到211,远远高于其他科技资源站点,分析其原因主要是由于数字化的科技图书文献的用户群较为广泛,其资源的需求程度和重复使用率相对较高,与其他类型科技资源相比具有不可比拟的优势。检测资源、医药卫生、地震、气象等科技资源与科研活动和人们的日常生活息息相关,由于其资源本身的使用范围较广,且适合通过信息化手段来进行此类资源的共享服务,因此这类科技资源站点的网络影响力较好。而农作物种质、水产种质资源等这类需要获得实物资源服务的资源站点,由于实物资源相对信息资源更为稀少、珍贵,使用该资源的也多为专业的科研人员为某些特定的科研活动所用,因此网络影响力也自然相对较低。

5 政策建议

为提升我国科技资源建设和综合利用水平,提高我国科技资源站点共享服务能力的和用户满意度,真正使科技资源站点成为科技工作者和科学家科研工具,建议从丰富资源站点内容建设、提高服务能力和增强互动性、个性化功能等方面提高资源站点的网络影响力。

(1)重视资源信息内容建设。从评价结果可以看出,科技资源站点的影响力评价结果是科技资源平台绩效评价的有力证明,但凡在科技资源建设(包括资源规模、质量)和开放共享服务(组织模式、运行机制)等方面建设比较好的平台,其资源站点不论是访问量、访问人数还是用户参与度都相对较高。资源站点内容建设一直是互联网行业发展的重要工作内容,对于提高用户的忠诚度和黏着度十分重要,如果科技资源站点能够长期提供丰富、独特、高质量的科技资源,必然会吸引大量的用户使用和分享。

(2)提高科技资源站点服务能力。提高科技资源站点访问量和服务能力是提高网站影响力的重要方法。科技资源站点与其他信息网站不同之处在于,科技资源站点是专业化的科技资源信息服务网站,不仅要提供及时、准确的资源信息服务,而且要提供快捷、便利的线下服务,例如实物资源的获取、大型仪器设备使用、资源用途和使用方法咨询等。如何使用户能够在海量科技资源信息中快速寻

找到所需资源,如何通过优化科技资源站点结构和用户使用界面,如何稳固一批专业化的服务人员队伍,是提高资源使用效率和网络影响力的关键。

(3)增强个性化、互动性功能。科技资源站点与通常的信息网站、商务网站不同,需要吸引、固定用户群,并且激发用户的科技创新意识。建议各类科技资源站点针对资源特定属性,加强资源深度挖掘,根据不同的用户需求形成专题服务产品,以提高用户参与度。鼓励用户对科技资源进行评价,借鉴最新的交互服务功能,开发第三方插件增强用户评论体验,以提高资源交互性和准确性。另外,还可以通过推广RSS订阅方式,帮助用户跟踪网站内容更新,对于长期使用资源站点用户来说,这个方法可以在未来带来持久的用户访问量,从而提高科技资源站点的网络影响力。

参考文献

- [1] 科技部计划司,财政部教科文司,国家科技基础条件平台中心.国家科技基础条件平台建设回顾与展望:整合、共享、创新[M].北京:中国科学技术出版社,2009.
- [2] 曾荷.电子政务信息资源的网络影响力评价研究[D].上海:华东师范大学,2007.
- [3] 段宁锋.网络链接分析与网站评价研究[M].北京:北京图书馆出版社,2005:70-71.
- [4] Most Influential Web Sites List Released[EB/OL]. [2004-01-05]. http://www.weboptimiser.com/search_engine_marketing_news/1707729.html.
- [5] 中国互联网络信息中心.中国互联网络网站影响力调查报告[R/OL]. [2004-01-05]. http://news.xinhuanet.com/ziliao/2003-01/21/content_699802.htm.
- [6] 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005.
- [7] 焦蓉.浅谈网站流量统计分析法[J].科技信息,2006(11):37.
- [8] 李长玲,王效岳,付鑫金.网站定量评价指标体系的构建与权值分配[J].图书情报工作,2008(7):52-55.
- [9] 张楠.如何进行网站运营数据分析[EB/OL]. [2010-01-17]. <http://www.chinaz.com/web/2009/1203/100077.shtml>.
- [10] 陈雅,郑建明.网站评价指标体系研究[J].中国图书馆学报,2002(5):57-60.